

PC기반의 easyBLAST 구현 Implementation of easy Local BLAST

김도완, 최한석*, 김동욱**
농업과학원 농업생명자원부, 목포대학교*,
(주)올비티**

Do-Wan Kim, Han Suk Choi*, Dong-Wook Kim**
National Institute of Agricultural Science,
Department of Computer Engineering Mokpo
National University*, AllDtec co. Ltd**

요약

본 논문에서는 BLAST 검색에 필요한 모든 절차들을 유전체연구자에게 익숙한 윈도우 환경에서 설계하고 구현하였다.

I. 서론

1990년대 인간 유전체 프로젝트(Human Genome Project)이후 염기나 단백질의 서열을 자동으로 분석할 수 있는 각종기기(DNA Sequencer, DNA Microarray 등)들의 획기적인 발전으로 인해 생물학적 정보의 양은 기하급수적으로 증가하고 있다. 이와같이 급증하고 있는 생물정보를 처리하기 위하여 생물정보학이 활발히 연구되고 있고 특히, 생물정보학(Bioinformatics) 연구에서 가장 많이 사용되고 있는 방법은 BLAST(Basic Local Alignment Search Tool)이다. 잘 모르고 있는 핵산(DNA 또는 RNA) 서열이나 단백질 서열이 주어지면, 먼저 분석하고자 하는 서열과 관계되는 서열들을 서열 데이터베이스(sequence database)로부터 찾아낸다. 그 다음에 데이터베이스에서 찾아진 서열들로부터 분석하고자 하는 서열의 성질들을 유추해낸다. 이러한 서열 데이터베이스 검색에서 가장 대중적으로 사용되고 있는 방법이 BLAST이며 BLAST는 대부분의 생물정보학 온라인 검색서버의 중심을 이루는 알고리즘이다. BLAST는 전체서열간의 최적 글로벌 정렬을 수행하는 것이 아니라 지역적 유사성(Local Similarity)이 있는 부분을 찾아 서열의 짝짓기 비교를 수행한다[1].

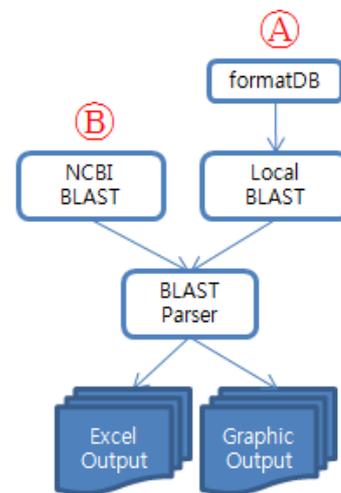
BLAST를 사용하는 방법에는 두 가지가 있다. 첫 번째는 미국 NCBI(National Center for Biotechnology Information) 사이트에서 웹을 통해서 BLAST 검색을 하는 것이다. 하지만 분석하고자 하는 서열들의 수가 아주 많을 때는 BLAST를 개인(또는 연구실) 컴퓨터에 설치하여 검색하는 것이 훨씬 효율적이다. 그러나 Local BLAST 검색을 하기 위해서는 유전체연구자들이 익숙하지 않은 DOS 또는 Linux 명령어를 습득해야 한다. 또한, BLAST 결과를 해석하기 위해 파싱을 해야 한다. 이러한 문제를 해결하기 위하여 본 시스템에서는 BLAST 검색에 필요한

모든 절차들을 유전체연구자에게 익숙한 윈도우 환경에서 설계하고 구현하였다.

II. 시스템 설계 및 구현

본 시스템에서는 BLAST를 윈도우 환경에서 사용하기 쉽도록 사용자관점에서 설계하였고 인터페이스 및 파서를 구현하였다.

1. 시스템 설계



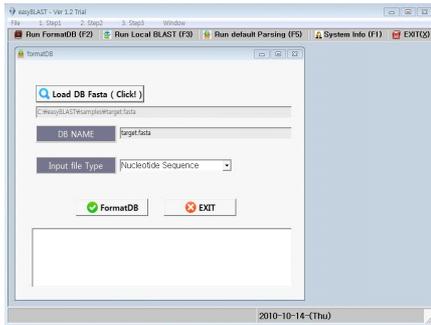
▶▶ 그림 1. easyBLAST 서비스 구조

본 시스템에서 설계한 easyBLAST는 두가지 서비스를 제공한다. (그림 1)에서 보듯이 ① Step은 Local BLAST를 사용하고자 할 때 과정을 포함하고 있고 ② Step은 easyBLAST에서 NCBI 홈페이지에 접근하지 않고도 NCBI에서 제공하는 DB와 BLAST 검색을 실행 할 수

있으며 BLAST 결과를 파싱하여 사용자가 편집이 가능한 excel 포맷과 그래픽 포맷으로 확인할 수 있도록 설계되었다.

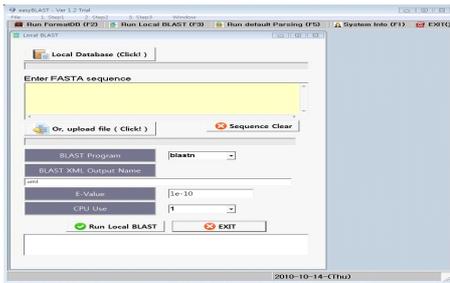
2. 시스템 구현

easyBLAST는 delphi7.0을 이용하여 구현하였으며 윈도우 환경에서 실행 되도록 개발하였다.



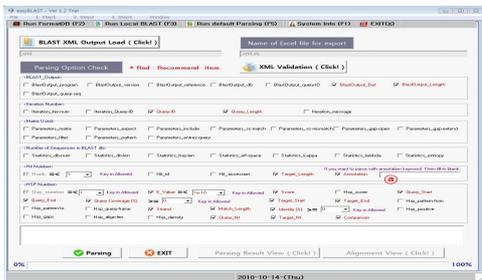
▶▶ 그림 2. formatDB

(그림 2)는 formatDB 메뉴를 보여주고 있으며 DNA 시퀀스 파일을 DB 파일 시스템으로 변환하여 준다. 특히, NCBI에서 다운받은 시퀀스 파일들은 gzip과 같은 압축파일 형태로 제공되어지나 easyBLAST에서는 이런 압축 파일도 자동으로 해제하여 DB를 생성할 수 있어 컴퓨터에 능숙하지 않는 사용자들도 쉽게 사용할 수 있도록 개발되었다.



▶▶ 그림 3. Local BLAST

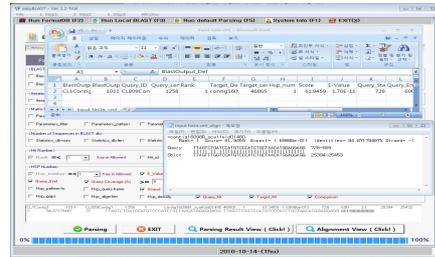
(그림 3)은 (그림 2)에서 생성한 DB에 사용자의 시퀀스를 BLAST할 수 있도록 구성되었으며 BLAST에 필요한 여러 옵션들을 직접 선택할 수 있도록 하여 사용자의 편의성을 증대하였다.



▶▶ 그림 4. easyBLAST 파서

(그림 4)는 BLAST결과로 나온 XML문서를 사용자가 원하는 결과로 나타낼 수 있도록 다양한 옵션값을 줄 수 있도록 구성되었다. 특히(그림 4)의 ④는 BLAST 결과 값 중 원하는 검색어를 포함하는 결과만을 파싱할 수 있는 검색기능을 제공하고 있다. 이러한 검색기능은 많은 연구자들이 필요한 기능이나 NCBI BLAST 검색등 다른 유사 프로그램에서는 제공하지 않고 있다.

(그림 5)는 파싱된 결과를 excel 파일과 서열정렬형태로 보여주고 있다. BLAST를 이용하는 연구자들에 따라 원하는 결과를 선택하여 볼 수 있도록 하였다.



▶▶ 그림 5. easyBLAST 파싱결과

II. 결론

본 시스템에서 구축된 easyBLAST 시스템을 이용하면 컴퓨터 지식이 없는 연구자들도 쉽게 BLAST 검색을 하고 결과를 제공 받을 수 있게 되며 서열 분석에 필요한 자료들을 구현된 시스템으로부터 빠르고 다양하게 제공 받을 수 있게 된다.

본 성과물은(논문, 산업재산권, 품종보호권 등)은 **농촌진흥청 포스트게놈다부처유전체연구사업 (PJ010455012017)의 지원에 의해 이루어진 것임**

■ 참고 문헌 ■

- [1] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool, J Mol Biol 1990, 215:403-410.